Verbs as the most "affective" words

Marina Sokolova and **Guy Lapalme** [1]

Abstract. We present a work in progress on machine learning of affect in human verbal communications. We identify semantic verb categories that capture essential properties when human communication combines spoken and written language properties. Information Extraction methods then are used to construct verb-based features that represent texts in machine learning experiments. Our empirical results show that verbs can provide a reliable accuracy in learning affect.

1 Introduction

In some social situations, there is a tendency to avoid adjectives and adverbs with explicit negative connotations. Their absence, or near absence, can create additional problems to Text Data Mining for automated and statistical learning of affect and emotions. We attribute this to the fact that negative adjective and adverbs discriminate more between positive and negative opinions than those with a positive affect [13]. In the absence of negative words in texts, the accuracy of affect and emotion classification usually declines. To overcome this problem, we have looked for other sets of features to represent texts in machine learning experiments not involving positive and negative words.

In this paper, we show that, under certain conditions, people's actions, expressed by verbs, allow an accurate machine learning of the conscious subjective aspect of feeling or emotion, i.e., affect. We apply Communication Theory to build semantic verb categories, then formalize their use by language patterns and apply Information Extraction to construct text features from them. Debates from the US Congress and consumer-written forum messages provide appropriate data for empirical support, because in both cases data contributors consciously state their feelings towards the discussed matters.

In the empirical part of the paper, we apply machine learning technique to the texts represented by the verb-based features by running regression and classification experiments. Regression problems of sentiment and emotion analysis have not been widely studied before as previous studies mainly focused on binary classification [10], sometimes solving a three-class classification problem [16]. Joined regression and classification learning allows a more detailed analysis of the applicability of our approach. In the absence of a direct affect labelling, we use given opinion labels as their estimates.

Our method does not rely on domain-specific and content words thus, it is applicable to study affect on data belonging to different domains. Our results can also be used to *recover* affect in situations in which participants do not give an explicit negative evaluation of the discussed matter.

[1] RALI, Département d'informatique et de recherche opérationnelle, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Québec, Canada, H3C 3J7, {sokolovm,lapalme}@iro.umontreal.ca

Category	Refers to	Examples
cognition	mental state	consider, hope, think, know
perception	activity of the senses	see,feel,hear
attitude	volition and feeling	enjoy,hate,love
activity	a continuing action	read, work, explain
event	happening or transition to another state	become,reply,pay,lose
process	continuing or eventual change of state	change,increase,grow

Table 1. The list of non-modal verb categories, their main semantic references, and examples of corresponding verbs.

2 Verb categories in human communication

Learning from records of human communication is one of the fastest growing areas of language and machine learning technologies. Such problems are more subjective and difficult to solve than traditional text classification and mining tasks [11], especially when the learning goal is the analysis of a communicated affect. They also require the development of methods to capture the relevant characteristics from a vast amount of data.

Stimulated by the English saying *"Actions speak louder than words"*, we looked at how verbs, which express actions in language, reveal a person's affect towards the discussed matter either emotionally or rationally. The emotional part may be expressed by attitude (enjoy, hate) and, partially, by the perception of the situation (smell, feel). The rational part may require the person to list facts such as events (meet, send) or the state of affairs (depend, have). To increase or diminish the communicative effect, people can use logic and politeness or imply possibility or necessity, that can be shown through the use of primary modals (can, will) or more conditional secondary modals (could, should) as was shown by the studies of Leech [8, 9].

We also consider that, under certain conditions, human communication combines characteristics of spoken and written communication. This happens when humans communicate through the Web or speak according to a prepared scenario such as in political debates. When such a situation occurs, we want to represent texts with the verbs that are most likely used in both spoken and written language.

For example, verbs denoting activity (play, write, send) and cognition verbs (think,believe) are the two most frequent categories when opinions are expressed in spoken-like language. Activity, the largest among verb categories, is the most frequent in all types of texts. Verbs denoting process (live, look, stay) often appear in written-like language, sometimes as often as activity verbs [2]. The high frequency of mental verbs is specific for spoken language [9, 14]. They are separated in three categories: attitude, perception and cognition. We defined the semantic categories, shown in Table 1, from verbs given in [8] to which we added their synonyms found the Roget's Interactive Thesaurus [1].

In Figure 1, we outline some involvement implications for patterns containing verbs. At the highest level, we consider whether the

closeness	$\longrightarrow$	*firstPerson* (*logic* | *physicalAction* | *mentalAction* | *state*)
distancing	$\longrightarrow$	`you` (*logic* | *physicalAction* | *mentalAction* | *state*)
logic	$\longrightarrow$	*primaryModal* | *secondaryModal*
physicalAction	$\longrightarrow$	[*modifier*] (*activity* | *event* | *process*)
mentalAction	$\longrightarrow$	[*modifier*] (*cognition* | *perception* | *attitude*)
state	$\longrightarrow$	[*modifier*] *havingBeing*
firstPerson	$\longrightarrow$	`I` | `we`
primaryModal	$\longrightarrow$	`can` | `may` | `will` | | `must` | ...
secondaryModal	$\longrightarrow$	`could` | `might` | `should` | `would`
activity	$\longrightarrow$	`read` | `work` | `explain` | ...
event	$\longrightarrow$	`become` | `reply` | `pay` | `send` | ...
process	$\longrightarrow$	`change` | `increase` | `stay` | ...
cognition	$\longrightarrow$	`believe` | `consider` | `hope` | ...
perception	$\longrightarrow$	`feel` | `hear` | `see` | `smell` | ...
attitude	$\longrightarrow$	`enjoy` | `fear` | `like` | `love` | ...
havingBeing	$\longrightarrow$	`have` | `be` | `depend` | `consist` | ...
modifier	$\longrightarrow$	*negation* | *adverb*

Figure 1. Rules (*non-terminal* $\longrightarrow$ alternative$_1$ | alternative$_2$ | ...) generalizing the use of verb categories. | separate alternatives, [] indicate optional parts and parenthesis are used for grouping. *non-terminal* must be replaced by one of the alternatives. Alternatives are composed of other non-terminals and `terminals` which are the pieces of the final string.

person involves herself in the statement (*firstPerson*) or projects on interlocutors (*you*):

closeness uses `I` or `we` to indicate a direct involvement of the author; sub-rules indicate different degrees of the author's involvement:

logic expresses permission, possibility, and necessity as the representation of logic, and superiority, politeness, tact, and irony as the representation of practice:

primaryModal such as can and may express direct possibility, permission or necessity of an action;

secondaryModal uses a more polite, indirect and conditional pattern than a primary modal and indicates more hypothetically and tentatively the author's intentions.

physicalAction denotes an author's goal-oriented actions (*activity*), actions that have a beginning and an end (*event*) and a series of steps towards a defined end (*process*). This pattern corresponds to a direct and active involvement of the author;

mentalAction uses mental action verbs, being more polite and tentative, that are a common face-saving technique and that mark openness for feedback;

state indicates personal characteristics and corresponds to actions without definite limits and strong differentiations.

distancing uses second person pronouns and shows how an author establishes distance from the matter.

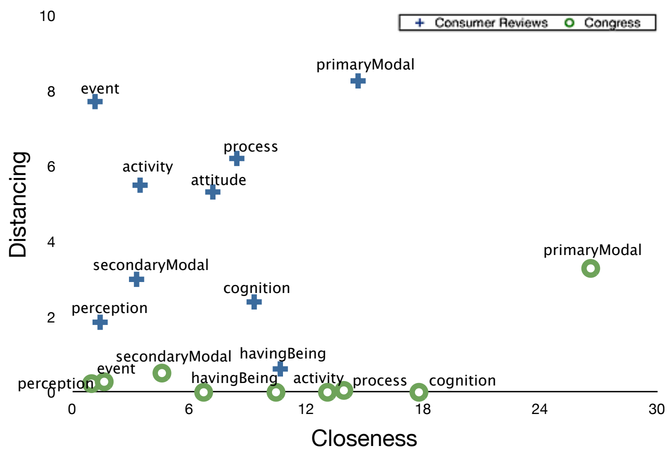

Figure 2. Distribution of verb categories in Congress debates and Consumer data.The horizontal axis estimates *closeness* (in per cent), the vertical axis – *distancing* (in per cent). Crosses denote Consumer reviews categories, circles – those in Congress debates. Labels indicate the verb categories of Figure 1.

3 Empirical support

In our empirical studies, we assume that positive and negative affect is revealed through the stated positive and negative opinion about the discussed matter. This is why we consider that opinion scores given in the corpus approximate the affect scores. We experimented with two kinds of data both combining spoken and written language properties: the first are consumer-written product reviews posted on the web which are loosely-edited, free structured texts, presumably written by the general population; the second are records of US Congress debates; they are more structured, edited and professionally written.

Consumer reviews introduced by Hu and Liu [4] in which text segments are manually tagged according to positive or negative opinions expressed by the reviewers, such as the following which is labelled +3 which means highly positive:

> this `is` *my* first digital camera , and what a ' toy ' *it* `is`! *i* `am` a software engineer and `am` very keen into technical details of everything *i* `buy`, *i* `spend` around 3 months before `buying` the digital camera; and *i* `must say`, g3 worth every single cent ...

To learn the strength of opinions, for the regression problem, we computed three numerical labels for each text: the number of positive tags, the number of negative tags, a signed sum of the two numbers. To solve classification problems, we applied unsupervised equal-frequency discretization to each of the numerical labels [3].

Congress debates We also used 1117 Congress debates [15] that either support or oppose a proposed legislature. Thomas et al. labeled texts by numerical polarity scores, computed by SUPPORT VECTOR MACHINE. SVM builds a decision surface that separates positive and negative texts. A score is the distance from a text to the surface. It can be positive or negative. The following excerpt has a positive score of 0.712:

> *we* `have known` that small businesses and working families `need` tax relief, and *we* `have fought` hard to `make` that happen so that *we* `see` the opportunity right there ...

For regression problems, we keep the scores as the data labels. For classification purposes, we use score signs as the data labels.

Verb distribution To illustrate and compare data's verb distributions, we calculated their frequencies and projected them with respect to *closeness* vs *distancing* axes as shown in Figure 2. As the resulting sets of points do not overlap, we conclude that the category distributions differ across these dimensions. The points of each set form a near-convex cluster with only one outlier: *havingBeing*, for consumer reviews and *primaryModal*, for Congress debates.

Information Extraction We constructed three feature sets based on the terminals of Figure 1:

1. The first feature set presents density and diversity of the words in each category. For a text T, for each verb category, we computed the number of word tokens and the number of word types for present, past and continuous forms. As a result, for each non-modal verb category we built six features. To represent modal verbs, we built four features, making 40 features in total.
2. The next set uses the 301 individual terminals as its features. Each terminal is represented by its occurrences in the text.
3. The third feature set expands the terminals with words occurring more than 5 times after or before a terminal.

See [12] for more information on the extraction process of verb information.

4 Learning experiments

We ran some learning algorithms available on Weka [17]. As mentioned in Section 3, we assumed that positive or negative subjective feeling is consciously revealed by the stated opinion. This assumption allowed us to use opinion labels as substitutes for the data affect labels.

Our first goal was to tackle regression (*quantitative*) learning problems. So far, machine learning experiments of sentiment analysis concentrated on classification (*qualitative*) tasks. Because of the novelty of this application, we wanted to try different types of algorithms to see what paradigms better learn the strength of the revealed affect. We chose KNN, a prototype-based algorithm, an optimization algorithm, SVM, and M5 TREES, a decision-based one. We applied BAGGING (bootstrap aggregating) to assess the influence of training data. In our experiments, BAGGING improved performance of M5 TREES, but not KNN nor SVM. We normalized each representation to eliminate the bias introduced by the text length.

Table 2 reports smallest relative absolute error RAE and corresponding root relative squared error $RRSQ$ obtained by the algorithms. The best performance, with the smallest error, was obtained on the Congress data. Positive consumer opinions were learned better than negative and overall opinions. An interesting phenomenon emerges when comparing algorithm performance – in terms of the learned correlation coefficients. The best performing algorithm in terms of accuracy is BAGGED M5 TREES. Since better accuracy implies that the algorithm learns dependencies between opinions and expressed actions better than other algorithms, we conclude that the output decision trees provide a reliable model of the data.

For Congressional debates, all output tree models agree that `demand`, `has` and `have` are the most important features, followed by `should` and `would`. Recall that we only report here the results of the best performing algorithms. Since this implies that the algorithms model better dependencies than other algorithms, we conclude that the strong language verbs have a positive correlation with attitude toward proposed legislations. On consumer review data, bagged trees placed `can`, `are` and `find` as the most important features for learning the overall opinions. Somewhat expectedly, `like` was among most decisive features for learning positive opinions. Learning negative opinions relied more on `be`, `am`, `would` and `should` than on other verbs.

To better display abilities of our approach, we performed a more traditional task of opinion classification. Again, we normalized each representation to eliminate the bias introduced by the text length. We chose SUPPORT VECTOR MACHINE (SVM) which is well-known for its high accuracy in text classification problems. Its use enabled us to compare our results with those of [15] obtained on the Congress debate data. They reported a test accuracy of 66.05 for positive/negative classification on the same data that we used for this work. The accuracy increased to 76.16 when they linked each data entry with previous speeches of the same speaker.

Our Congress results (78.14) have a better accuracy, even though we did not use previous records of speakers or other data reinforcements; the results are reported in the right part of Table 3. These results thus show that the *expressed actions do speak loud*. Under certain conditions, they reveal more than the previous history of the same speaker. For consumer reviews, learning positive opinions was easier than learning negative and overall opinions. Our method's accuracy is close to human-human agreement on positive and negative sentiments, when it is based on verbs [5]. More details on learning with verb-based features are provided in [12].

5 Related work

Sentiment analysis focuses on whether a text, or a term is subjective, bears positive or negative opinion or expresses the strength of opinion. Application of learning algorithms - through classification - has been pioneered by Lee et all [10]. However, Lee and many authors that followed her, used machine learning algorithms on reviews written by only four professional critics. This means that the algorithms were trained and tested on overly specific undiversified data. To achieve a comparable accuracy on the Congress data, they had to enhance data with previous speeches of speakers. Our goal is to seek general enough methods that can work with an unrestricted number of data contributors.

For automating recognition and evaluation of the expressed opinion, texts are represented through N-grams or patterns and then classified as opinion/non-opinion, positive/negative, etc. [6]. Syntactic and semantic features that express the intensity of terms are used to classify opinion intensity [16]. These works do not consider a hierarchy of opinion disclosure. We, however, built a pragmatic-lexical hierarchy of the use of semantic categories that allows us to interpret machine learning models formulated in lexical items and in terms of the pragmatics.

Various verb semantic classification schemes have been suggested and used for different purposes. Biber et al [2] examine word distribution, lexico-grammatical patterns and grammatical/discourse factors of four text genres: conversation records, fiction, news and academic writing. The authors suggest seven verb categories: activity, mental, communication, existence, occurrence, causative, aspectual. We think that these verb categories are not specific enough to distinguish between the verb's use in communicating personal opinions and other texts. We opted to build verb categories that reflect peculiarities of expressing personal opinions.

VerbNet [7] assigns verbs to 57 lexical-semantic categories such as *urge* (ask,persuade), *order* (command,require), *wish* (hope, expect), *approve* (accept,object). Since this work do not consider whether texts exhibit communication characteristics, the verb categories the

Algorithms	Consumer reviews						Congress	
	positive		negative		overall		debates	
	RAE	$RRSE$	RAE	$RRSE$	RAE	$RRSE$	RAE	$RRSE$
KNN	91.19	87.97	90.77	88.70	93.56	96.50	78.74	86.60
SVM	80.98	84.15	89.33	96.71	91.38	94.38	90.89	94.80
BM5P	80.26	82.21	87.21	85.81	89.82	96.61	73.73	78.84

Table 2. Smallest $RelativeAbsoluteError$ and $RootRelativeSquaredError$ obtained by the algorithms. Rows report results for each algorithm. Columns report results for each problem. For each problem, the smallest RAE is in $italic$.

Features	Consumer reviews						Congress	
	positive		negative		overall		debates	
	Acc	$Recall$	Acc	$Recall$	Acc	$Recall$	Acc	$Recall$
Categories	74.52	74.50	63.64	61.50	66.24	67.30	65.70	67.90
Terminals	76.12	75.80	66.56	67.20	70.06	74.50	69.63	72.00
Terminals-B	76.43	75.70	67.83	73.20	73.60	75.20	70.61	73.40
Collocations	77.75	79.00	68.33	69.50	73.82	78.90	75.18	77.60
Collocations-B	**78.87**	80.10	70.95	71.40	75.21	79.70	**78.14**	81.10

Table 3. Accuracy and corresponding true positive rates obtained by SVM. Rows report results for each feature set. Columns report results for each problem. For each problem, the largest accuracy is reported in **bold**. Baselines are the majority class accuracy: for the consumer data – 52.22, for Congress – 59.76.

authors suggest do not capture specifics of communication. We focused on verb categories in communicative texts, in which speakers communicate their opinions about the discussed matters.

6 Conclusion and future work

In this study, we have shown the importance of relations between expressed actions and affect. We formalized expressed actions by building language patterns of modal, event, activity, process, cognition, perception, state verbs and personal pronouns. We applied machine learning methods to establish quantitative relations between the use of verb categories and affect.

Our use of regression and classification methods allows to perform a more detailed learning than previous studies that usually defined their problems either as binary classification or multi-class classification problems. On two data sets, consumer reviews [4] and the US Congress debates [15], we showed that regression problems were successfully learned by BAGGED M5 TREES, whereas SVM obtained a reliable accuracy in classification problems. Our method extracts all its information from only the given data. Other methods could only achieve a similar accuracy by adding personal information about speakers, such as the history of previous comments [15]. However, such type of additional information is not often easily available.

Learning affect from the used verbs becomes practically justified and, indeed, desirable when a social context dictates avoidance of negative adjectives and adverbs, because empirical results showed that negative adjective and adverbs discriminate better between positive and negative emotions than positive ones. In the future, we intend to analyze the use of different types of verb modifiers (always, never). We are also interested in learning the correspondence between a revealed affect and pragmatics of communication, e.g. intensity and immediacy. Another venue for future work is to investigate the phenomenon of impression building, i.e. how texts allow inference of an author's abilities or intentions.

Acknowledgements

This work has been supported by *Natural Sciences and Engineering Research Council* of Canada. The authors thank Fabrizio Gotti for technical support of experiments.

REFERENCES

[1] Roget's interactive thesaurus, 2006. http://thesaurus.reference.com/.
[2] D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan, *Longman Grammar of Spoken and Written English*, Longman, 1999.
[3] M. Boulle, 'Optimal bin number for equal frequency discretizations in supervized learning', *Intelligent Data Analysis*, **9**(2), 175–188, (2005).
[4] M. Hu and B. Liu, 'Mining opinion features in customer reviews', in *Proceedings of Nineteeth National Conference on Artificial Intelligence (AAAI-2004)*. AAAI Press, (2004).
[5] S.-M. Kim and E. Hovy, 'Determining the sentiment of opinions', in *Proceedings of the of the 20th international conference on Computational Linguistics (COLING - 2004)*, pp. 1367–1373, (2004).
[6] S.-M. Kim and E. Hovy, 'Crystal: Analyzing predictive opinions on the web', in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 1056–1064, (2007).
[7] K. Kipper, A. Korhonen, N. Ryant, and M. Palmer, 'Extending verbnet with novel verb classes', in *Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, pp. 25–32, (2006).
[8] G. Leech, *Meaning and the English Verb*, Longman, 2004.
[9] G. Leech and J. Svartvik, *A Communicative Grammar of English*, Longman, third edn., 2002.
[10] B. Pang, L. Lee, and S. Vaithyanathan, 'Thumbs up? sentiment classification using machine learning techniques', in *Proc Empirical Methods of Natural Language Processing EMNLP'02*, pp. 79–86, (2002).
[11] M. Sokolova and G. Lapalme, 'Performance measures in classification of human communication', in *Proceedings of the 20th Canadian Conference on Artificial Intelligence (AI'2007)*, pp. 159 – 170. Springer, (2007).
[12] M. Sokolova and G. Lapalme. Do actions speak loud? semantic verb categories in expression of opinions, 2008. submitted elsewhere.
[13] M. Sokolova and G. Lapalme. A simple and effective information extraction method for opinion analysis, 2008. submitted elsewhere.
[14] M. Sokolova and S. Szpakowicz, 'Language patterns in the learning of strategies from negotiation texts', in *Proceedings of the 19th Canadian Conference on Artificial Intelligence (AI'2006)*, pp. 288 – 299. Springer, (2006).
[15] M. Thomas, B. Pang, and L. Lee, 'Get out the vote: Determining support or opposition from congressional floor-debate transcripts', in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 327–335, (2006).
[16] T. Wilson, J. Wiebe, and R. Hwa, 'Recognizing strong and weak opinion clauses', *Computational Intelligence*, **22**(2), 7399, (2006).
[17] I. Witten and E. Frank, *Data Mining*, Morgan Kaufmann, 2005.